

Visualizações Gráficas Iniciais

Transcrição

[00:00] Agora que a gente já fez alguns tratamentos na base de dados, alguns ajustes, a gente fez algumas caracterizações na base, a gente categorizou, ajustou alguns campos. Vamos entender melhor as variáveis que estão na base de dados.

[00:15] Então, eu vou mostrar para vocês como faz alguns gráficos e algumas porcentagens, já para a gente entender um pouco melhor da estrutura. Então, eu vou abrir aqui um novo chunk aqui, para a gente já ir escrevendo. Então, eu vou colocar aqui o chunk. Primeiramente, então, para que a gente precisa fazer exatamente?

[00:42] O que que a gente vai fazer? A gente precisa entender a frequência absoluta relativa de alguns campos, como é que a... quando... principalmente para os campos que são fatores. Então, quando a gente tem um campo numérico, a gente pode fazer uma média, variância, achar a mediana.

[01:00] Então, a gente tem alguns cálculos matemáticos que a gente pode fazer para entrar a média. Agora, quando a gente tem campos, como a gente tem aqui, por exemplo, estado civil, região, como é que a gente resume isso, como é que a gente (vem). Então a gente precisa entender, fazer algumas frequências.

[01:16] Então, vamos primeiramente... aqui a gente vai observar frequências absolutas e relativas desses campos. Então, como é que a gente pode fazer isso aqui? Como é que a gente vai fazer? Eu vou criar uma nova... a gente cria uma nova variável, [XXIninteligívelXX] criar frequência absoluta, eu vou criar essa variável chamada frequência absoluta.

[01:57] E ela vai receber aqui, geralmente, o que que a gente vai fazer? A gente vai fazer um table, a gente vai olhar um table, chama o nome do arquivo que a gente está trabalhando, que aqui é o funcionário já tratados e a variável de interesse que a gente quer fazer.

[02:14] Vamos olhar o estado civil primeiro. Então, eu vou fazer essa frequência absoluta do campo estado civil. Se a gente pegar e rodar só essa linha aqui, vou dar o "Ctrl + Enter", ele já cravou e se eu pegar e executar só a frequência absoluta aqui, o que que ele tem? Está vendo?

[02:36] Então, já tem aqui solteiro 16, então tem 16 casos, 16 pessoas são solteiras e 19 pessoas são casadas. Ok, já tem a frequência absoluta, da mesma forma, eu vou calcular agora a frequência relativa. Então eu crio essa nova variável chamada frequência relativa, que ela vai receber também, agora sim é uma proporção, o próprio ponto table, de quem?

[03:09] Da minha frequência absoluta, que eu acabei de criar. Então, se eu executar, eu vou executar só essa linha, se eu executar só a frequência, vou dar o "Ctrl + Enter" aqui em cima da frequência relativa, ele calculou a frequência relativa. Então, eu tenho 45% da minha base, são solteiras, então pessoas solteiras e 54% são pessoas casadas.

[03:36] Eu posso querer agora, só que perceba, eu fiz essas duas variáveis aqui de forma separadas, eu posso até... eu vou copiar e colar aqui, vou executar de novo ele vai chamar a própria variável. Então, eu ele já calculou aqui, mas eu poderia estar agrupando isso com uma única tabela.

[03:59] Então, para agrupar, eu vou fazer da seguinte forma, eu crio uma nova... pelo comando rbind, eu vou poder juntar essas duas variáveis aqui, o comando o rbind, ele junta variáveis. Então, eu vou ter a minha frequência absoluta, virgula frequência relativa, que eu acabei de criar.

[04:24] Agora, assim, se eu... eu vou apagar essas duas soltas aqui, se eu executar essa linha aqui, ela vai juntar as duas tabelas, é como se eu tivesse [XXIninteligívelXX], então ficou mais visível aqui, aparentemente. Então, eu tenho aqui...

eu fiz uma tabela cruzada.

[04:41] Então, eu tenho 16 pessoas na minha base que são solteiras, frequência relativa é 0.45, eu poderia pegar essa frequência relativa aqui e multiplicar por 100, até para visualizar um pouco melhor. Se eu multiplicar ela por 100 aqui, então fiz a mesma coisa. Então, eu tenho aqui 45.7, 54.28.

[05:09] Então, só para visualizar um pouco melhor essa parte de... já observar a base de dados. Poderia fazer isso aqui para outros campos, só mudando exatamente a minha variável de interesse aqui, como a gente viu aqui, estado civil, poderia fazer a mesma coisa para as demais variáveis, outras variáveis aqui, por exemplo, região.

[05:41] Vamos pegar a próxima aqui, por exemplo, região. Deu, simplesmente, agora que a estrutura está criada, eu venho aqui e mudo a minha variável aqui, vou dar um... vou copiar e colar aqui para a gente ficar mais... só para mostrar, que aí, depois é bem fácil criar outras... fazer essa análise.

[06:00] Então, invés de estado civil aqui, eu mudo a variável, vou chamar agora, região, a minha variável região e aí, vou colocar isso aqui num outro chunk, só para ficar mais organizado. Então, aqui em baixo, região. Aí, eu executo, tenho aqui a minha frequência de como é que está observada a região.

[06:30] Então, eu tenho aqui capital, interior e outros, que não tem... apareceu com frequência zero. Eu tenho 10 pessoas que estão com a capital, 10 funcionários que estão na capital, 12 funcionários que estão no interior, com 45% e 54%. Agora, eu vou mostrar aqui, como é que faz a visualização de alguns gráficos, como é que a gente poderia criar esses... visualização gráfica.

[07:11] A visualização gráfica no R, existem alguns pacotes mais sofisticados, como o ggplot, então tem uma infinidade de plots, mas eu vou mostrar os mais simples aqui, depois, se alguém quiser, tiver mais interesse, o pessoal pode pesquisar por outros pacotes mais avançados, uma visualização... quanto mais sofisticado o seu gráfico, mais comando ele vai ter, mais atributos que você vai ter que colocar no gráfico.

[07:40] No R, você praticamente, literalmente você constrói o gráfico mesmo. Então, o primeiro gráfico que a gente vai fazer aqui é o gráfico de barras, então vamos fazer o gráfico de barras, como é que a gente vai fazer esse gráfico de barras? O gráfico de barras no R mais simples que tem, ele é chamado de barplot.

[08:01] Então, o que que eu vou fazer? Eu vou criar uma variável chama bp, que vai receber esse barplot, esse barplot, ele vai ser a minha table, vai ser a table, eu estou procurando a tabela, func_t, que é funcionários tratados e vou pegar a minha variável estado civil, o meu estado civil aqui.

[08:33] Aí, eu defino... preciso definir o col, que é o colnames, o colnames não, seria a definição da cor dessas barras, geralmente eu defino aqui com um vetor de cores. Então, eu vou colocar aqui blue e red, o mais simples aqui. Então, eu criei essa variável barplot, com isso, eu vou executar.

[09:08] Então, ele cria um gráfico bem simples aqui de barras. Agora, um outro tipo de gráfico que eu vou mostrar para vocês é o gráfico de pizza, bastante comum também, bem, assim, bastante comum, principalmente para exemplificar campos que são fatores, não numéricos, para exemplificar justamente categorias, gráfico de pizza.

[09:42] Gráfico de pizza, como é que a gente vai fazer? Eu também preciso... primeiramente, eu preciso definir aqui quais são os rótulos, então eu vou chamar de rótulos. Esses rótulos, eles vão ter as definições, como se fosse um label, então eu vou ter dois rótulos aqui.

[10:01] Opa, cria um vetor também de rótulos, o primeiro aqui vai ser o solteiro e o segundo casado. Aí, eu preciso definir aqui esses rótulos também, eles vão receber aqui... eu vou precisar definir ele como... vou acrescentar no gráfico, então eu vou colar aqui no gráfico os rótulos em formato de percent.

[10:49] Então, estou definindo aqui que eu quero esse gráfico - eu vou subir aqui um pouquinho aqui, só para visualizar melhor - em formato de porcentagem e... de tal forma que eu tenha aqui também, de novo, vou colocar agora... eu vou atribuir o que seria essa junção, também dando um paste aqui, colando para dentro da junção do que?

[11:20] Do rótulos em porcentagem. Eu estou escrevendo exatamente o que eu vou querer em porcentagem e um separador aqui, que é espaço. Defini quais os parâmetros que eu quero no meu gráfico e aí, por fim, eu tenho o meu pie, que que seria pizzas, seria formado pelas fatias.

[11:53] Vamos executar antes de fazer isso aqui, vamos executar só essa parte aqui, o que que ele faz? Ele não fez nada. Opa, escrevi errado aqui, rótulos, vou executar de novo. Por enquanto ele não fez nada, por quê? Porque eu preciso definir o tipo de gráfico aqui mesmo, que seria o que?

[12:10] As minhas fatias, o meu label, seriam os labels que são... quem são os meus labels? São exatamente os rótulos que eu defini e as... conjunto de cores agora, conjunto de cores que é o meu vetor também... eu vou colocar aqui blue e red. Vou executar e aí, formou o gráfico de pizza aqui.

[12:50] Veja que ele é bem... tem que ir definindo mesmo as coisas, ir construindo. Aqui eu construí em um.... saiu em output, aqui em baixo um output. Tem um comando no R bastante... que eu poderia estar colocando um do lado do outro, mas se eu quiser colocar um do lado do outro, é como como se eu estivesse fazendo uma matriz com duas colunas.

[13:12] Então, tem um comando aqui que possibilita isso aqui, que esse comando chama par mfrow e aí, eu defino como é que eu quero dividir a minha tela para colocar esse gráficos, eu vou colocar da seguinte forma, uma linha, duas colunas, par mfrow. Eu coloco isso aqui, antes de fazer os gráficos.

[13:51] Então, eu vou juntar esses dois gráficos aqui num único chunk, eu vou colocar... estar escrevendo aqui, eu vou estar juntando os gráficos, só para ficar... Então, eu vou pegar o meu primeiro gráfico de pizza, vou pegar o comando aqui, que a gente fez aqui.

[14:12] Então, vou escrever num outro chunk, joga aqui logo abaixo e pego os meus comandos aqui do gráfico de pizza e joga no próximo comando aqui do chunk. Vou executar tudo, o que que ele vai fazer? Ele vai dividir a minha tela e vai colocar um do lado do outro, só mudando a apresentação.

[14:33] Então, está vendo? Ficou mais... um pouco mais apresentável, poderia fazer como fiz aqui, um sobre o outro, mas também, se a gente quiser, a gente pode colocar um do lado do outro. Aí, a gente usa esse comando par mfrow aqui, divide a tela, poderia estar dividindo em mais, sempre vai dividir como linha e coluna, como se fosse uma matriz.

[14:55] Era isso que eu queria mostrar aqui para vocês, essa visualização, então, como fazer gráficos tipo barra, pizza, para ir entendendo exatamente melhor e a parte também de porcentagem aqui do nosso banco de dados.

[15:10] Então, a gente pode fazer... exatamente fazer essa frequência absoluta, relativa ou fazer a visualização de alguns gráficos para entender exatamente um pouco mais e resumir a nossa análise.

[15:23] Então, era isso que eu tinha para mostrar aí.